Compressibility of Mixed-State Signals

Masato Koashi and Nobuyuki Imoto
CREST Research Team for Interacting Carrier Electronics, School of Advanced Sciences, The Graduate University for Advanced Studies (SOKEN), Hayama, Kanagawa, 240-0193, Japan

Abstract

We present a formula that determines the optimal number of qubits per message that allows asymptotically faithful compression of the quantum information carried by an ensemble of mixed states. The set of mixed states determines a decomposition of the Hilbert space into the redundant part and the irreducible part. After removing the redundancy, the optimal compression rate is shown to be given by the von Neumann entropy of the reduced ensemble.

PACS numbers:03.67.-a, 03.67.Hk

Consider a source that generates a message i with probability p_i. Sequences of the messages independently drawn from this source can be compressed into sequences of bits and decompressed back to the original sequences of messages. The necessary and sufficient number of bits per message allowing asymptotically faithful compression and decompression is given by the Shannon entropy $S = -\sum_i p_i \log_2 p_i$. This result, called the noiseless coding theorem [1], is one of the core results of the classical information theory. The quantum analogue of this theorem, which will naturally form a basis of quantum information theory, was first considered by Schumacher [2]. In this quantum data compression, the source emits a system in a quantum state ρ_i with probability p_i, and sequences of the systems emitted from this source are assumed to be compressed into qubits. It was shown [2–4] that when all ρ_i are pure, the least number of qubits allowing asymptotically faithful recovery of the original states is given by the von Neumann entropy $S(\rho) = -\mathrm{Tr}\rho\log_2\rho$ of the density operator of the ensemble $\rho = \sum_i p_i\rho_i$. When $\{\rho_i\}$ includes mixed states, the problem is still open. Since compression schemes applicable to the pure-state signals can also be successfully used for the mixed-state cases [5], the optimal compression rate I_{p} is bounded from above by the von Neumann entropy, namely, $I_{\mathrm{p}} \leq S(\rho)$. It has also been proved [6] that the Levitin-Holevo function [7], $I_{\mathrm{LH}} = S(\rho) - \sum_i p_i S(\rho_i)$, is a lower bound for I_{p}, namely, $I_{\mathrm{LH}} \leq I_{\mathrm{p}}$.

The aim of this Letter is to identify the optimal compression rate for the mixed-state ensemble $\mathcal{E} = \{p_i, \rho_i\}$. We first introduce a function $I_{\mathrm{R}}(\mathcal{E})$ that is given as the von Neumann entropy of a reduced ensemble $\mathcal{E}_{\mathrm{R}} = \{p_i, \sigma_i\}$. The ensemble $\mathcal{E}_{\mathrm{R}}$ is derived from $\mathcal{E}$ by stripping off the redundant parts. Then we prove that $I_{\mathrm{R}}(\mathcal{E})$ is equal to the optimal compression rate I_{p}.

The problem considered here is formulated as follows. Suppose that the source produces the ensemble $\mathcal{E} = \{p_i, \rho_i\}$, namely, it emits a system in a quantum state ρ_i with probability p_i. Using this source N times, we obtain a state $\rho_\lambda^N \equiv \rho_{i_1} \otimes \cdots \otimes \rho_{i_N}$ acting on a Hilbert space $\mathcal{H}^N \equiv \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_N$ with probability $p_\lambda^N = p_{i_1} \ldots p_{i_N}$, where λ represents a set of indexes $\{i_1, \ldots, i_N\}$. We assume that the dimension d of each space $\mathcal{H}_n$ is finite. Now $\mathcal{H}^N$ is given to Alice, who compresses the signal ρ_λ^N into $\tilde{\rho}_\lambda$ acting on a Hilbert space $\mathcal{H}_{\mathrm{C}}$ with a dimension usually smaller than Nd. This process is generally written by a quantum operation (linear completely positive trace-preserving map) $\rho_\lambda^N \to \tilde{\rho}_\lambda = \Lambda_{\mathrm{A}}(\rho_\lambda^N)$. The operation Λ_{A} is independent of λ since only the systems $\mathcal{H}^N$ are given to Alice and no additional information on λ is available. The coded signal $\tilde{\rho}_\lambda$ is passed on to Bob through a noiseless channel, and he decompresses the signal by a quantum operation $\tilde{\rho}_\lambda \to \rho'_\lambda = \Lambda_{\mathrm{B}}(\tilde{\rho}_\lambda)$, where ρ'_λ acts on $\mathcal{H}^N$. To measure the quality of the whole process $\rho_\lambda^N \to \rho'_\lambda$, we use the fidelity F [8] given by $F(\rho,\sigma) \equiv [\mathrm{Tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}]^2$. The quality of a compression scheme specified by $(\Lambda_{\mathrm{A}}, \Lambda_{\mathrm{B}})$ for the ensemble $\mathcal{E}$ is given by the average fidelity

$$\bar{F} \equiv \sum_\lambda p_\lambda F(\rho_\lambda^N, \rho'_\lambda). \tag{1}$$

Now, for a fixed source $\mathcal{E}$, consider a sequence of compression schemes $(\Lambda_{\mathrm{A}}^{(N)}, \Lambda_{\mathrm{B}}^{(N)})$ with increasing N. When $\lim_{N\to\infty}\bar{F} = 1$, the sequence gives asymptotically faithful compression of $\mathcal{E}$. Such sequences are called *protocols* [6]. For a given protocol P, the quantity $R(P)$ characterizing the asymptotic degree of compression is defined through the size of $\mathcal{H}_{\mathrm{C}}$ measured in the number of qubits, namely,

$$R(P) \equiv \lim_{N\to\infty}(\log_2 \dim\mathcal{H}_{\mathrm{C}})/N. \tag{2}$$

Then, the optimal compression rate $I_{\rm p}(\mathcal{E})$ for the ensemble $\mathcal{E}$ is formally defined as

$$I_{\rm p}(\mathcal{E}) \equiv \inf_{P} R(P). \tag{3}$$

This means that for arbitrary small $\delta > 0$, asymptotically faithful compression is possible if $I_{\rm p} + \delta$ qubits per message is given, and it is impossible if $I_{\rm p} - \delta$ qubits per message is given.

A useful tool used for stripping off the redundant part in $\mathcal{E}$ and deriving the formula for $I_{\rm p}(\mathcal{E})$ below is the theory [9] that characterizes the quantum operations which preserves a set of states $\{\rho_i\}$ (maps ρ_i to ρ_i) acting on a Hilbert space $\mathcal{H}$. To state the results of this theory, it is convenient to express quantum operations in unitary representation, namely, by unitary operations U acting on the combined space $\mathcal{H} \otimes \mathcal{H}_{\rm E}$, where $\mathcal{H}_{\rm E}$ represents an auxiliary system initially prepared in a standard pure state $\Sigma_{\rm E}$. Then, it was shown [9] that, given $\{\rho_i\}$, we can find a decomposition of $\mathcal{H}_{\rm A}$ defined as the support of $\sum_i \rho_i$ ($\mathcal{H}_{\rm A}$ is generally a subspace of $\mathcal{H}$) written as

$$\mathcal{H}_{\rm A} = \bigoplus_l \mathcal{H}_{\rm J}^{(l)} \otimes \mathcal{H}_{\rm K}^{(l)}, \tag{4}$$

in such a way that any U preserving $\{\rho_i\}$ is expressed in the following form

$$U(\mathbf{1}_{\rm A} \otimes \Sigma_{\rm E}) = \bigoplus_l \mathbf{1}_{\rm J}^{(l)} \otimes U_{\rm KE}^{(l)}(\mathbf{1}_{\rm K}^{(l)} \otimes \Sigma_{\rm E}), \tag{5}$$

where $U_{\rm KE}^{(l)}$ are unitary operators acting on the combined space $\mathcal{H}_{\rm K}^{(l)} \otimes \mathcal{H}_{\rm E}$. Under this decomposition, ρ_i is written as

$$\rho_i = \bigoplus_l q^{(i,l)} \rho_{\rm J}^{(i,l)} \otimes \rho_{\rm K}^{(l)}, \tag{6}$$

where $\rho_{\rm J}^{(i,l)}$ and $\rho_{\rm K}^{(l)}$ are normalized density operators acting on $\mathcal{H}_{\rm J}^{(l)}$ and $\mathcal{H}_{\rm K}^{(l)}$, respectively, and $q^{(i,l)}$ is the probability for the state to be in the subspace $\mathcal{H}_{\rm J}^{(l)} \otimes \mathcal{H}_{\rm K}^{(l)}$. $\rho_{\rm K}^{(l)}$ is independent of i, and $\{\rho_{\rm J}^{(1,l)}, \rho_{\rm J}^{(2,l)}, \ldots\}$ cannot be expressed in a simultaneously block-diagonalized form. An explicit procedure to obtain this particular decomposition is also given in [9].

The form of Eq. (6) implies that the spaces $\mathcal{H}_{\rm K}^{(l)}$ are redundant in the ensemble $\mathcal{E} = \{p_i, \rho_i\}$. Consider the states $\sigma_i \equiv \bigoplus_l q^{(i,l)} \rho_{\rm J}^{(i,l)}$ in which the redundancy has been removed, and let $\mathcal{E}_{\rm R} \equiv \{p_i, \sigma_i\}$ be the corresponding ensemble. The von Neumann entropy of $\mathcal{E}_{\rm R}$ can be regarded as a function of the ensemble $\mathcal{E}$, denoted as $I_{\rm R}(\mathcal{E})$, since the decomposition (6) is determined by the set $\{\rho_i\}$. What we prove below is that the optimal compression rate $I_{\rm p}(\mathcal{E})$ is given by the function $I_{\rm R}(\mathcal{E})$.

We begin the proof by noting that the two ensembles $\mathcal{E}$ and $\mathcal{E}_{\rm R}$ are completely interchangeable, namely, there exist quantum operations $\Lambda_{\sigma\rho}$ and $\Lambda_{\rho\sigma}$ that satisfy $\Lambda_{\sigma\rho}(\rho_\lambda^N) = \sigma_\lambda^N$ and $\Lambda_{\rho\sigma}(\sigma_\lambda^N) = \rho_\lambda^N$. If a compression scheme $(\Lambda_{\rm A}, \Lambda_{\rm B})$ for ρ_λ^N is given, we can compose a compression scheme $(\Lambda_{\rm A}\Lambda_{\rho\sigma}, \Lambda_{\sigma\rho}\Lambda_{\rm B})$ for σ_λ^N. Since the fidelity does not decrease under a quantum operation [10,11], we have the inequality $F(\rho_\lambda^N, \Lambda_{\rm B}\Lambda_{\rm A}(\rho_\lambda^N)) \leq F(\sigma_\lambda^N, \Lambda_{\sigma\rho}\Lambda_{\rm B}\Lambda_{\rm A}\Lambda_{\rho\sigma}(\sigma_\lambda^N))$. Hence the composed scheme always has a better or equal average fidelity. This implies that if a protocol for $\mathcal{E}$ with an asymptotic degree of compression R is given, we can compose a protocol for $\mathcal{E}_{\rm R}$ with the same degree R [6,10]. Consequently, we have $I_{\rm p}(\mathcal{E}) \geq I_{\rm p}(\mathcal{E}_{\rm R})$. Since a similar argument can be made with ρ and σ interchanged, we obtain the equality

$$I_{\rm p}(\mathcal{E}) = I_{\rm p}(\mathcal{E}_{\rm R}). \tag{7}$$

Now it is suffice to consider the cases where $\{\rho_i\}$ have no redundancy, namely, $\sigma_i = \rho_i$ and $\mathcal{E}_{\rm R} = \mathcal{E}$, and we will prove the relation $I_{\rm p}(\mathcal{E}) = S(\rho)$ in these cases. Since we already have the inequality $I_{\rm p}(\mathcal{E}) \leq S(\rho)$, what we need is the opposite inequality, $I_{\rm p}(\mathcal{E}) \geq S(\rho)$. We will give a sketch of the proof first.

In a compression-decompression scheme $(\Lambda_{\rm A}, \Lambda_{\rm B})$, the state eventually evolves as $\rho_\lambda^N \to \rho'_\lambda = \Lambda(\rho_\lambda^N)$, where $\Lambda \equiv \Lambda_{\rm B}\Lambda_{\rm A}$. In this process, the marginal state in the first system ($\mathcal{H}_1$) evolves from ρ_{i_1} to ${\rm Tr}_{2\ldots N}(\rho'_\lambda)$. This evolution can be regarded as a result of a quantum operation Λ_1, defined as

$$\begin{aligned}\Lambda_1(\rho_i) &\equiv \sum p_{i_2}\ldots p_{i_N} {\rm Tr}_{2\ldots N}\Lambda(\rho_i \otimes \rho_{i_2} \otimes \cdots \otimes \rho_{i_N}) \\ &= {\rm Tr}_{2\ldots N}\Lambda(\rho_i \otimes \rho \otimes \cdots \otimes \rho).\end{aligned} \tag{8}$$

Note that Λ_1 is determined by Λ and the *total* density operators (ρ) of the initial state ensembles of the other $N-1$ systems. In a protocol, a scheme $(\Lambda_{\rm A}, \Lambda_{\rm B})$ for large N is nearly perfect. For this scheme, Λ_1 will almost preserve the states $\{\rho_i\}$. The decomposition (4) for $\{\rho_i\}$ satisfying $\rho_i = \sigma_i$ can be simplified as $\mathcal{H}_{\rm A} = \bigoplus_l \mathcal{H}_{\rm J}^{(l)}$ since $\mathcal{H}_{\rm K}^{(l)}$ is a one-dimensional space. Correspondingly, the requirement (5) for preserving $\{\rho_i\}$ can be written as

$$U(\mathbf{1}_{\rm A} \otimes \Sigma_{\rm E}) = \bigoplus_l \mathbf{1}_{\rm J}^{(l)} \otimes U_{\rm E}^{(l)}\Sigma_{\rm E}, \tag{9}$$

where $U_{\rm E}^{(l)}$ are unitary operators acting on $\mathcal{H}_{\rm E}$. The operation Λ_1, which nearly preserves $\{\rho_i\}$, should thus be approximately written in the form (9). Next, take a diagonalization of the total density operator, $\rho = \sum_{ls} p_{l,s}|l,s\rangle\langle l,s|$, in such a way that for a fixed l, the set $\{|l,s\rangle\}$ forms a basis of $\mathcal{H}_{\rm J}^{(l)}$. Let us consider an ensemble $\mathcal{E}_\perp \equiv \{p_{l,s}, \rho_{l,s} \equiv |l,s\rangle\langle l,s|\}$ composed of orthogonal pure states. If we replace the source from $\mathcal{E}$ to $\mathcal{E}_\perp$ in the scheme $(\Lambda_{\rm A}, \Lambda_{\rm B})$, the operation Λ_1 does not change because the total density operator is identical for the two ensembles. Then, the error in the transmission of $|l,s\rangle$ will be small since the operation of the form (9) preserves $\{|l,s\rangle\}$. This means that by a projection measurement in the basis $\{|l,s\rangle\}$, classical information close to $NS(\rho)$ bits can be sent through the channel $\mathcal{H}_{\rm C}$. This implies $\log_2 {\rm dim}\mathcal{H}_{\rm C} \gtrsim NS(\rho)$. Combined with the definitions (2) and (3), we have $I_{\rm p}(\mathcal{E}) \gtrsim S(\rho)$.

The strict proof is given by clarifying the meaning of 'nearly' in the above sketch, by introducing several measures (f and g below) characterizing the nearness. In unitary representation, any quantum operation for the system $\mathcal{H}_1$ can be represented by a unitary operator U in $d + d^2 \equiv n$ dimension [12], acting on the combined space of $\mathcal{H}_1$ and an auxiliary system $\mathcal{H}_{\rm E}$ with dimension d^2. Let us introduce two nonnegative continuous functions $f, g : U(n) \to R$ that measure how $U \in U(n)$ is close to the form (9). The first one is defined as $f(U) \equiv 1 - \sum_i p_i F(\rho_i, \Lambda_U(\rho_i))$, where $\Lambda_U(\rho_i) \equiv {\rm Tr}_{\rm E}[U(\rho_i \otimes \Sigma_{\rm E})U^\dagger]$. Since $f(U) = 0$ iff $\Lambda_U(\rho_i) = \rho_i$ for all i, $f^{-1}(0)$ is equal to the set of U that can be expressed in the form (9). The other measure is related to the average error probability of the transmission of $\mathcal{E}_\perp$, defined as $p_{\rm e} \equiv 1 - \sum_{l,s} p_{l,s}{\rm Tr}(\rho_{l,s}\Lambda_U(\rho_{l,s}))$. For later convenience, we use the function $g(U)$ defined through $p_{\rm e}$, namely, $g(U) \equiv H(p_{\rm e}) + p_{\rm e}\log_2(d-1)$ with $H(p) \equiv -p\log_2 p - (1-p)\log_2(1-p)$. Since the form (9) preserves $\{\rho_{l,s}\}$, $g(U)$ is zero for any $U \in f^{-1}(0)$. An important relation

between the two measures is that if g is away from zero, f must also be away from zero. This is proved as follows. Let us define the set $\bar{X}_\delta \equiv \{U|g(U) \geq \delta\}$ for arbitrary $\delta > 0$. Since g is continuous, $\bar{X}_\delta$ is a closed subset of $U(n)$. Since $U(n)$ is compact and f is continuous, the image $f(\bar{X}_\delta)$ is closed in R. $\bar{X}_\delta \cap f^{-1}(0) = \emptyset$ implies that $0 \notin f(\bar{X}_\delta)$. Therefore, $f(\bar{X}_\delta)$ has its minimum $\eta(\delta) > 0$. This result will be used to derive the inequality (11) below. Note that the functional dependence of η on δ is determined by $\mathcal{E}$, and is independent of N.

Next, we consider the transmission of classical variable $\{(l,s)\}$ through the source $\mathcal{E}_\perp$ and the scheme $(\Lambda_{\rm A}, \Lambda_{\rm B})$. Let $X_k (k = 1, \ldots, N)$ be independent random (vector) variables with $\Pr\{X_k = (l,s)\} = p_{l,s}$, and $X \equiv \{X_1, \ldots, X_N\}$. Suppose that the value of X_k is encoded to the state $|l,s\rangle$ in the system $\mathcal{H}_k$, the compression-decompression scheme is applied to combined system $\mathcal{H}^N$, and finally the state in each $\mathcal{H}_k$ is measured by the projection to the basis $|l,s\rangle$, producing a result Y_k. The transmitted data is represented by $Y \equiv \{Y_1, \ldots, Y_N\}$. The quantum operation Λ_k on each system $\mathcal{H}_k$ can be written in a similar form as (8). Let us take a unitary representation $U_k \in U(n)$ for Λ_k. A lower bound for the mutual information $I(X;Y) \equiv H(X) - H(X|Y)$ in this example is obtained as follows. Since X_k are independent, we have $H(X) = \sum_k H(X_k) = NS(\rho)$. From the general properties of entropy, we obtain the following inequalities [13]: $g(U_k) \geq H(X_k|Y_k)$ (Fano's inequality), $H(X_k|Y_k) \geq H(X_k|Y)$ (conditioning reduces entropy), and $\sum_k H(X_k|Y) \geq H(X|Y)$ (independence bound on entropy). Combining these, we have $I(X;Y) \geq NS(\rho) - \sum_k g(U_k)$. On the other hand, $I(X;Y)$ cannot exceed the capacity of the channel $\mathcal{H}_{\rm C}$, namely, $\log_2 {\rm dim}\mathcal{H}_{\rm C} \geq I(X;Y)$. We thus arrive at the relation

$$\sum_k g(U_k)/N \geq S(\rho) - (\log_2 {\rm dim}\mathcal{H}_{\rm C})/N. \tag{10}$$

Now let us suppose that the number of available qubits per message is smaller than $S(\rho)$, namely, $(\log_2 {\rm dim}\mathcal{H}_{\rm C})/N = S(\rho) - \delta$ with $\delta > 0$. Since the numbering of the systems $\mathcal{H}_k$ is arbitrary, we can generally assume that $g(U_1)$ is not smaller than any other $g(U_k)$. Then, from the relation (10) we have $g(U_1) \geq \delta$, or equivalently, $U_1 \in \bar{X}_\delta$. As shown above, this implies $f(U_1) = 1 - \sum_i p_i F(\rho_i, \Lambda_1(\rho_i)) \geq \eta(\delta) > 0$. From the properties of the fidelity function F, we obtain

$$\begin{aligned}
\bar{F} &= \sum_\lambda p_\lambda^N F(\rho_\lambda^N, \Lambda(\rho_\lambda^N)) \\
&\leq \sum_\lambda p_\lambda^N F(\rho_{i_1}, {\rm Tr}_{2\ldots N}\Lambda(\rho_{i_1} \otimes \rho_{i_2} \otimes \cdots \otimes \rho_{i_N})) \\
&\leq \sum_i p_i F(\rho_i, \Lambda_1(\rho_i)) \leq 1 - \eta(\delta)
\end{aligned} \tag{11}$$

since the fidelity does not decrease under partial trace (the first inequality) and $F(\sigma, \rho)$ is convex as a function of ρ (the second). The average fidelity of the compression-decompression scheme never exceeds $1 - \eta(\delta) < 1$ for any N, where $\eta(\delta)$ is independent of N. This means that no protocols exist that satisfy $R(P) = S(\rho) - \delta$. Hence $I_{\rm p}(\mathcal{E}) \geq S(\rho)$. Combined with the opposite inequality $I_{\rm p}(\mathcal{E}) \leq S(\rho)$, we obtain $I_{\rm p}(\mathcal{E}) = S(\rho)$ for the ensemble $\mathcal{E}$ satisfying $\mathcal{E} = \mathcal{E}_{\rm R}$. Together with Eq. (7), we obtain the formula for general $\mathcal{E}$,

$$I_{\rm p}(\mathcal{E}) = I_{\rm R}(\mathcal{E}), \tag{12}$$

which is the main result of this Letter. For convenience, we repeat the definition of the function $I_{\rm R}(\mathcal{E})$: From $\mathcal{E} = \{p_i, \rho_i\}$, determine $\mathcal{E}_{\rm R} = \{p_i, \sigma_i\}$ through the decomposition (6). Then, $I_{\rm R} = S(\sigma)$ with $\sigma = \sum_i p_i \sigma_i$.

The protocols we considered above is asymptotically reversible, namely, Bob is required to asymptotically reproduce everything that was given to Alice. Bob can thus compress the reproduced signals again with the same degree of compression. This class of protocols is called *blind* protocols, and the obtained bound $I_{\rm p}$ is called *passive information* [6,10]. In another scenario, not only the system $\mathcal{H}^N$ but also the identity of the state ρ_λ^N, namely, the index λ is disclosed to Alice. Bob still has to decompress the signal without additional knowledge of λ. This class of protocols is called *visible* protocols, and the corresponding optimal compression rate is called *effective information* $I_{\rm eff}$ [6,10]. This scheme is irreversible and cannot be repeated, but the compression rate $I_{\rm eff}$ may be better than $I_{\rm p}$. The difference $I_{\rm d} \equiv I_{\rm p} - I_{\rm eff}$ is called *information defect*. For an ensemble of pure states, it was shown that the information defect is zero [4]. While the identity of $I_{\rm eff}$ is still an open question, the derived form of $I_{\rm p}$ assures the presence of nonzero information defect for an ensemble of mixed states, which can be shown as follows. In the second scenario, Alice can compress the classical value λ into the length of Shannon entropy, and send it directly to Bob. This indicates $I_{\rm eff} \leq -\sum_i p_i \log_2 p_i$. For example, if $p_1 = p_2 = 1/2$, $I_{\rm eff} \leq 1$. On the other hand, by allowing the dimension d large, we can find examples of ρ_1 and ρ_2 with arbitrarily large $I_{\rm p}$, according to the result (12).

Finally, we would like to raise several problems which is worthy of future investigation. What we have proved in this Letter corresponds to the so-called the weak converse of Shannon's noiseless coding theorem, namely, if $I_{\rm p} - \delta$ qubits are available per system, the fidelity cannot reach unity in $N \to \infty$. For classical or pure-state ensembles, the strong converse holds, namely, the fidelity goes to zero when $N \to \infty$. Whether this statement holds for mixed-state cases or not is an important open question. In the proof of the main result, we utilized the observation that any protocols for an mixed-state ensemble $\mathcal{E}$ with no redundancy ($\mathcal{E} = \mathcal{E}_{\rm R}$) can be used to transmit the 'purified' ensemble $\mathcal{E}_\perp$ with errors asymptotically negligible *per message*. The requirement ($\bar{F} \to 1$) for the compression protocols for $\mathcal{E}_\perp$ is more stringent, namely, the total errors for the whole N messages must be negligible. Whether the protocols for $\mathcal{E}$ always works as compression protocols for $\mathcal{E}_\perp$ or not is another open question.

In summary, we derived the formula for the optimal compression rate (passive information) for a general mixed-state ensemble $\{p_i, \rho_i\}$. This will give a measure of how much information is stored in the ensemble of quantum states in terms of qubits. We have also shown the presence of nonzero information defect, namely, there are cases where knowing the identity of states gives more efficient compression.

This work was supported by a Grant-in-Aid for Encouragement of Young Scientists (Grant No. 12740243) and a Grant-in-Aid for Scientific Research (B) (Grant No. 12440111) by Japan Society of the Promotion of Science.

REFERENCES

[1] E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
[2] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).
[3] R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).
[4] H. Barnum, C. A. Fuchs, R. Jozsa, and B. Schumacher, Phys. Rev. A **54**, 4707 (1996).
[5] H. -K. Lo, Opt. Commun. **119**, 552 (1995).
[6] M. Horodecki, Phys. Rev. A **57**, 3364 (1999).
[7] A. S. Holevo, Probl. Peredachi Inf. **8**, 63 (1973).
[8] R. Jozsa , J. Mod. Opt. **41**, 2315 (1994).
[9] M. Koashi and N. Imoto, `arXiv quant-ph/0101144`.
[10] M. Horodecki, Phys. Rev. A **61**, 052309 (2000).
[11] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, Phys. Rev. Lett. **76**, 2818 (1996).
[12] B. Schumacher, `arXiv quant-ph/9604023`.
[13] T. M. Cover and J. A. Thomas, *Elements of information theory* (John Wiley & Sons, New York, 1991), chap. 2.